

JONATHAN CACHAT PHD

Dir of MLops | Sr Data Scientist & Data Engineer | Neuroscience (PhD)

614-982-0521 cell @jc@jcatchat.com



HIGHLIGHTS

- 20+ years of research & data-driven engineering experience. paradigms, research questions, experimental designs, hypothesis testing, statistical evaluation, visualization and communication.
- 15+ years working with various data types and standard technical tools. My current daily drivers are GCP, Python, RStudio & RapidMiner. Building acquisition and analysis pipelines for streaming IoT data, Structured/Unstructured databases, Time-Series data (GBs), Movement Pattern data (TBs), Video/Visual data (TBs), Behavioral & Genetic data (TBs), and data scraped, crawled, pushed, pulled, batched.
- 10+ years of entrepreneurial leadership, raised \$3M, lead small PhD teams & large business teams, delivered business process optimization and automation. Sought out subject matter expert and consultant, delivering results for clients across the board.
- 10+ years in the Cannabis space – especially pushing data-driven strategies in each vertical.

SKILL STACK

Google Cloud Platform

BigQuery (columnar)

BigTable (key-value, wide-column)

Cloud Storage (blob)

Cloud Composer (Apache Airflow & DAGs)

DataFlow (Apache Beam)

DataPlex (Data Quality Monitoring)

Vertex AI (MLops & Data Science)

Geospatial Analytics & Movement Pattern Analysis

Generative AI / Embeddings, Vector-stores & LLM training

Python / SQL / LLMs

pandas, polars, numpy

langchain & llamaindex

PROFESSIONAL EXPERIENCE

Data Scientist & Data Engineer

CCV Research

2015 - Present Remote (Ohio)

provide operational & technology expertise for business process improvement, focused on R&D to enable data-driven, rapid decisions - getting insights from data faster. Reduce COGs to maximize competitive advantages, expand products & human activity patterns to improve recommendation and semantic profiling models. Modernize & Automate business processes in the cloud to rapidly scale services, minimize infrastructural risks & drive revenue. dozens of happy clients.

- Tools: Google Cloud Platform (GCP) - BigQuery, BigTable, Composer (Airflow), Dataflow (Beam), Compute, Cloud Functions, Cloud Storage, Pub/Sub, VertexAI. advanced SQL, Python (polars, data-prep, langchain). & GenAI skills
- Areas: Finance, AdTech, Energy, HealthCare, Defense, Geospatial
- Several production-deployed LLM chatbots/agents using LangChain, LlamaIndex, Vector databases - typically embedded into BigQuery via VertexAI/gemini or other fine-tuned model.

Dir of Data Science & MLops

Torchlight AI

2022 - 2024 Remote (US & EU)

lead DS, DnA & Engineering teams through complete cloud infra transformation on time & under budget. oversaw, managed and orchestrated ingestion of over 100MM rows of new data nightly. pre-processed global geolocation data, netflow & user activity logs to stage for analytics. feature engineering, model evaluation & containerized deployment for analytics to extract patterns of life. Migrated pipelines from GCS, BigQuery & Airflow to BigTable, DataFlow (Beam), Spanner & VertexAI.

- behavioral intelligence platform for analytics & models of emerging risk, fraud & other target activities of interest. government, defense & private markets
- Analyzed global human movement patterns in near-real-time, covering {1,000+} datasets daily.
- custom model development for dept of defense partners
- generative ai for internal dev tools & user experience improvements

Sr Data Engineer (AI/ML pipelines)

Cardinal Health

2021 - 2023 Remote (Worldwide)

Worked with Pharma Commercial Enablement, Digital Solutions team at Cardinal Health & to understand stakeholder's business challenges & develop analytical, cloud-based solutions that deliver regular business insights. excelled by performing intital EDA on raw data, developing ingestion methods into Warehouses/Lakes, & pre-processing to stage for feature engineering, training, model evaluation & containerized deployment.

- Healthcare Services, 14th highest revenue generating US company
- \$16M in Type 1 benefit over last 3 years
- Delivery of critical work like infrastructure modernization and pipeline refactoring to optimize management & overhead costs

Analytics Engineer & Architect

Seinergy

2019 - 2020 Remote (US & Canada)

- transitioned the team, data, and business workflows over to the Google Cloud Platform and continue to optimize and expand these systems to meet ongoing business needs.
- The transition to GCP allowed Seinergy to save thousands of dollars, expand the business productivity, and boost revenues within 7 months.
- As a Technical Director with Zativa, coordinated with Ohio State University (OSU) and the OH Dept of Agriculture to plant & monitor the first legal hemp growing season at the Wooster Extension Campus.

Scientific Director | Laboratory Sciences

Midway Labs & Hocking College Cannabis Lab

2017 - 2021 Ohio

Cannabis Analytical Testing Labs for Ohio Medical Cannabis Program

- Achieved the opening of {2} state-licensed cannabis analytical labs in Ohio MMCP by leading application development and operational strategy.
- Pioneered the nation's 1st accredited Associate's Degree in Cannabis Lab Technician, impacting {100+} students.

SKILL STACK

Methods / Models

Geospatial Analytics & Movement Pattern Analysis

Generative AI / Embeddings, Vector-stores & LLM training

PROJECTS & PORTFOLIOS

GitHub

github.com/cachatj

Work Portfolio

blog.jcachat.com/portfolio

ORCID

<https://orcid.org/0000-0002-6504-0869>

Google Scholar

https://scholar.google.com/citations?user=PYWxq_AAAAAJ&hl=en

PROFESSIONAL EXPERIENCE

co-founder, Chief Executive & Technology Officer

[Conscious Cannabis Ventures](#), [Fleurish Farms](#), [SunGrown Zero](#)

📅 2014 - 2018 📍 Sonoma County, California

emerging cannabis industry start-ups - management, branding & cultivation R&D. raised \$6MM venture capital funding with team of PhDs

- Developed Novel Hybrid-Light Cultivation Facility through targeted R&D integrating grow room intelligence (Grownetics), responsive LEDs, and Solatube Daylighting Devices.
- Reduced energy and water use by > 75%, matched production rates, and improved product quality as defined by potency and terpene profiles.
- [SunGrown Indoor: Can Sunlight from Solatube Daylight Harvesting Systems Support Indoor Cannabis Growth?](#)

Neuroinformatics Associate & Open Data Evangelist

[Neuroscience Information Framework](#)

📅 2010 - 2013 📍 San Diego, California

Graduate Research Manager

[Kalueff Lab, Dept of Pharmacology](#)

📅 2009 - 2012 📍 New Orleans, LA

EDUCATION

Post-doctoral Fellow

[University of California, Davis \(UCD\)](#)

📅 2014

PhD & MS, Neuroscience

[Tulane University](#)

📅 2012

- Neuroscience; Psychopharmacology Research (behavior models, movement activity profiles)
- 2 books, 50+ peer-reviewed articles - [google scholar](#)
- [Psychedelic Zebrafish: the effects of LSD and MDMA on zebrafish behavior and physiology](#)

BS, Social Neuroscience

[Denison University](#)

📅 01/2004 - 12/2008

Social Neuroscience

[University of Chicago](#)

📅 01/2006 - 12/2006

Certification, Digital Media Arts

[Northwestern University](#)

📅 01/2004 - 12/2004